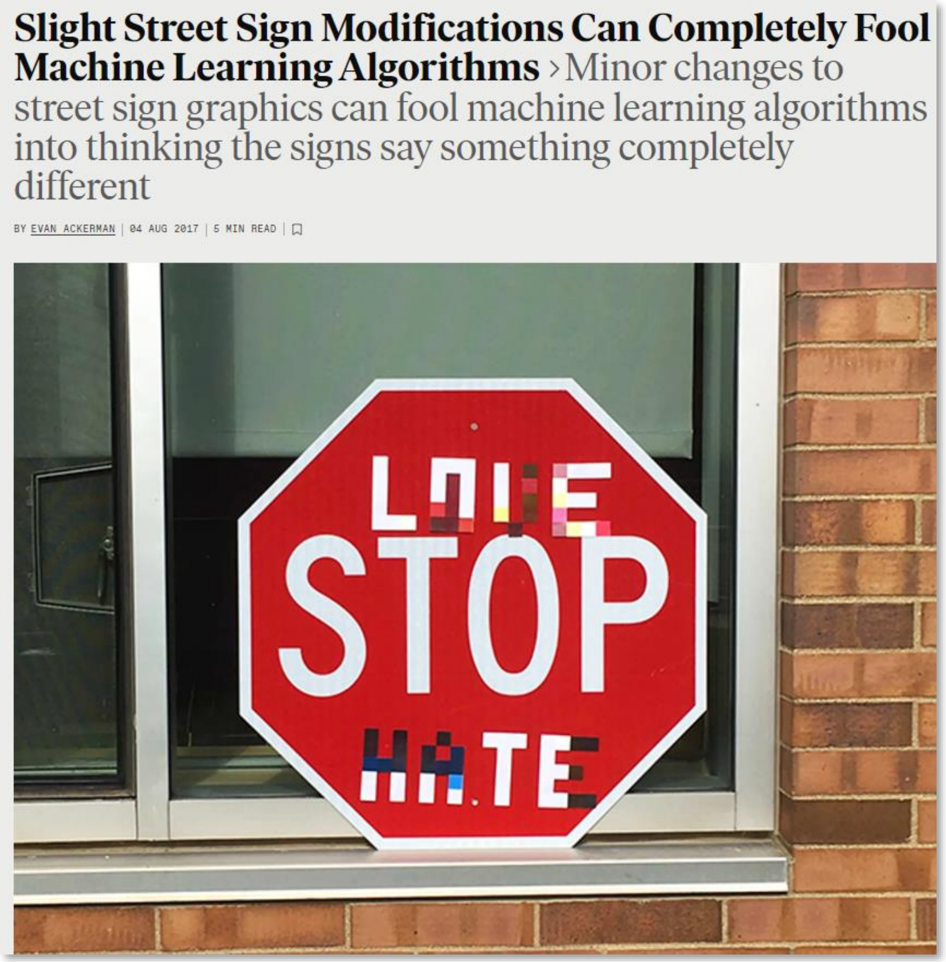
Robustness Evaluation of Visual Perception Systems

Huang Yihao

1, Sep, 2023

Background

- **Self-driving faces serious security problem**


Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms › Minor changes to street sign graphics can fool machine learning algorithms into thinking the signs say something completely different

BY EVAN ACKERMAN | 04 AUG 2017 | 5 MIN READ


BBC Sign in Home News Sport Reel Worklife

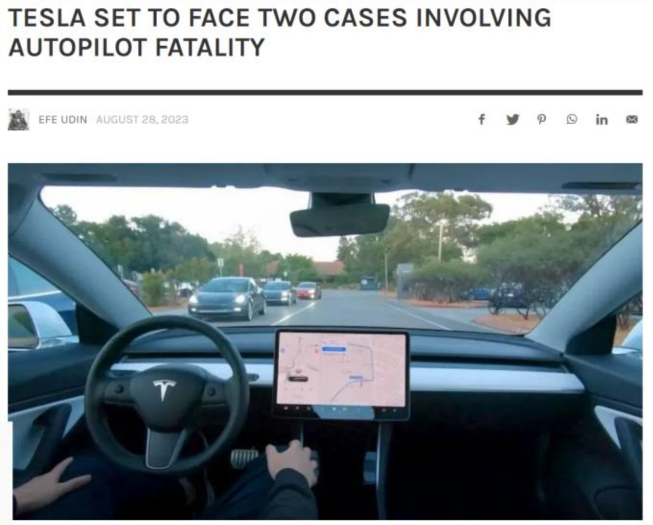
NEWS

Home | War in Ukraine | Coronavirus | Climate | Video | World | Asia | UK | Business | Tech | Science

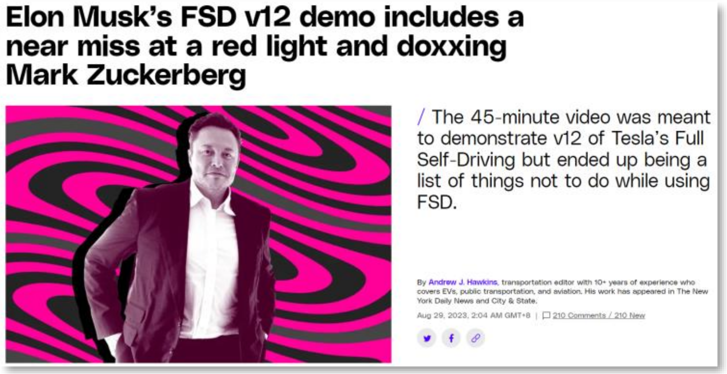
Tech

Uber's self-driving operator charged over fatal crash

16 September 2020


TESLA SET TO FACE TWO CASES INVOLVING AUTOPILOT FATALITY

EFE UDIN AUGUST 28, 2023


Elon Musk's FSD v12 demo includes a near miss at a red light and doxxing Mark Zuckerberg

/ The 45-minute video was meant to demonstrate v12 of Tesla's Full Self-Driving but ended up being a list of things not to do while using FSD.

Background

- **Visual perception systems**

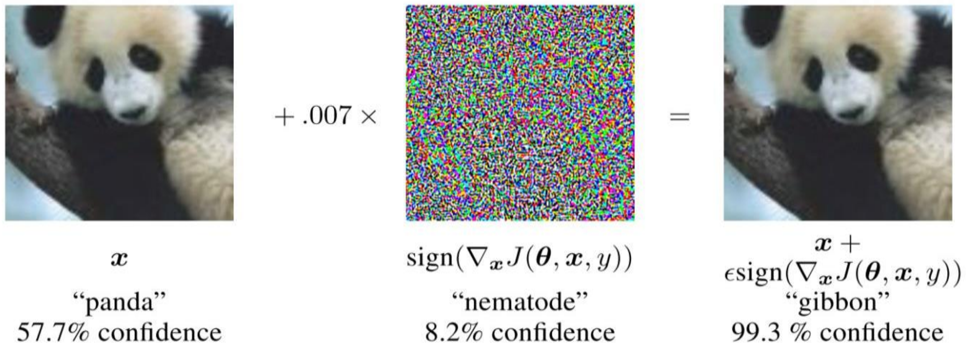
Segmentation

Recognition

Detection

- **Security problem**

Adversarial attack



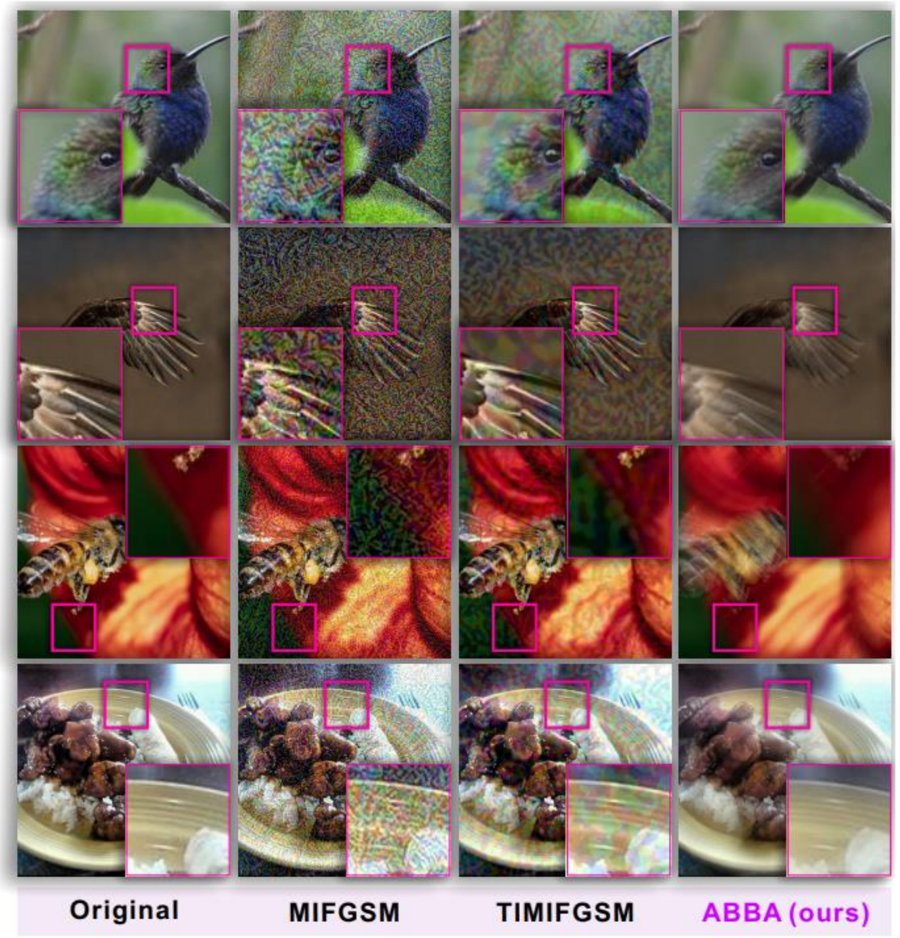
Corruption

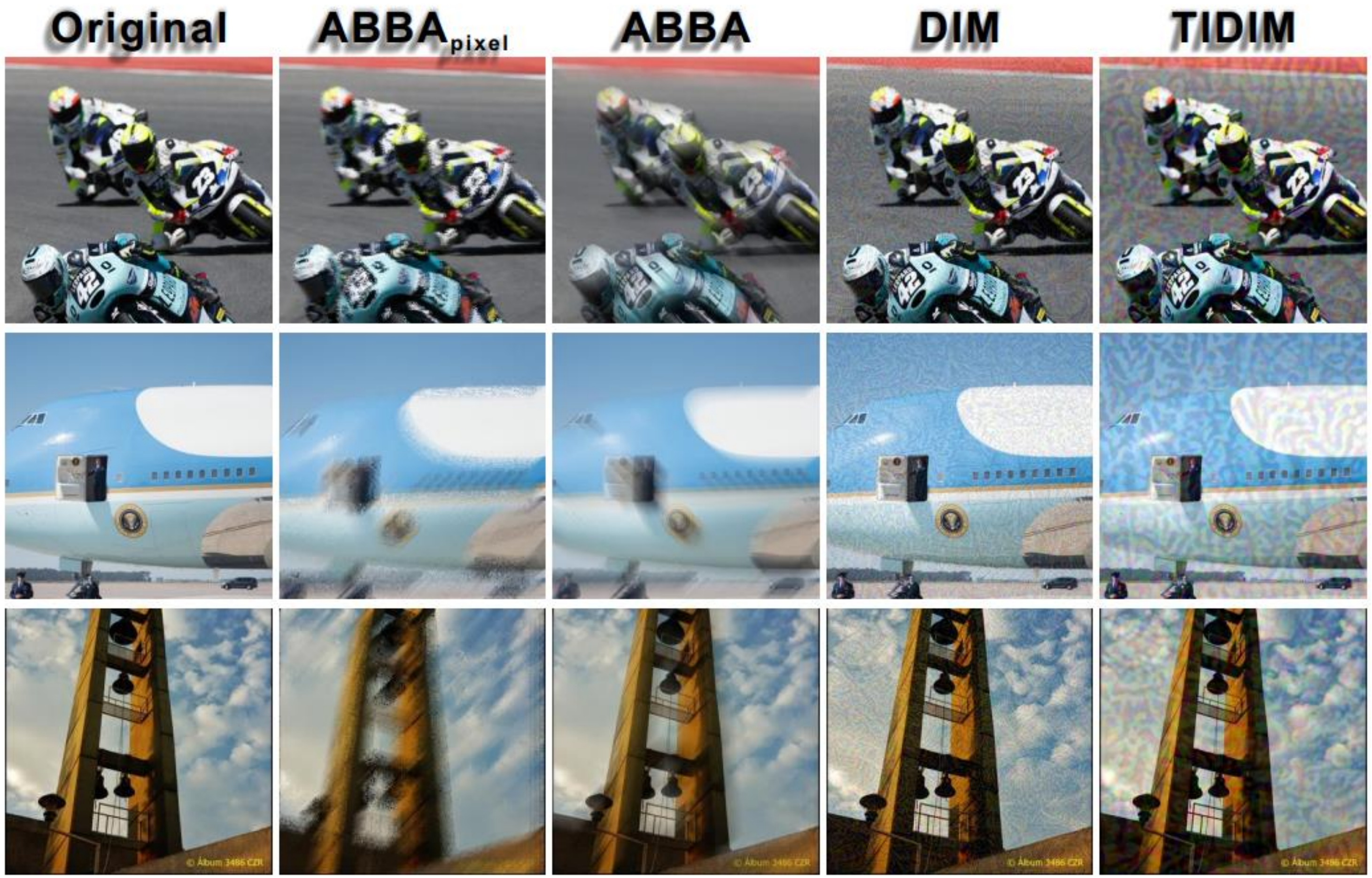
Dark

Snow

Rain

Attack





Guo Q, Juefei-Xu F, Xie X, et al. Watch out! motion is blurring the vision of your deep neural networks[J]. NeurIPS, 2020

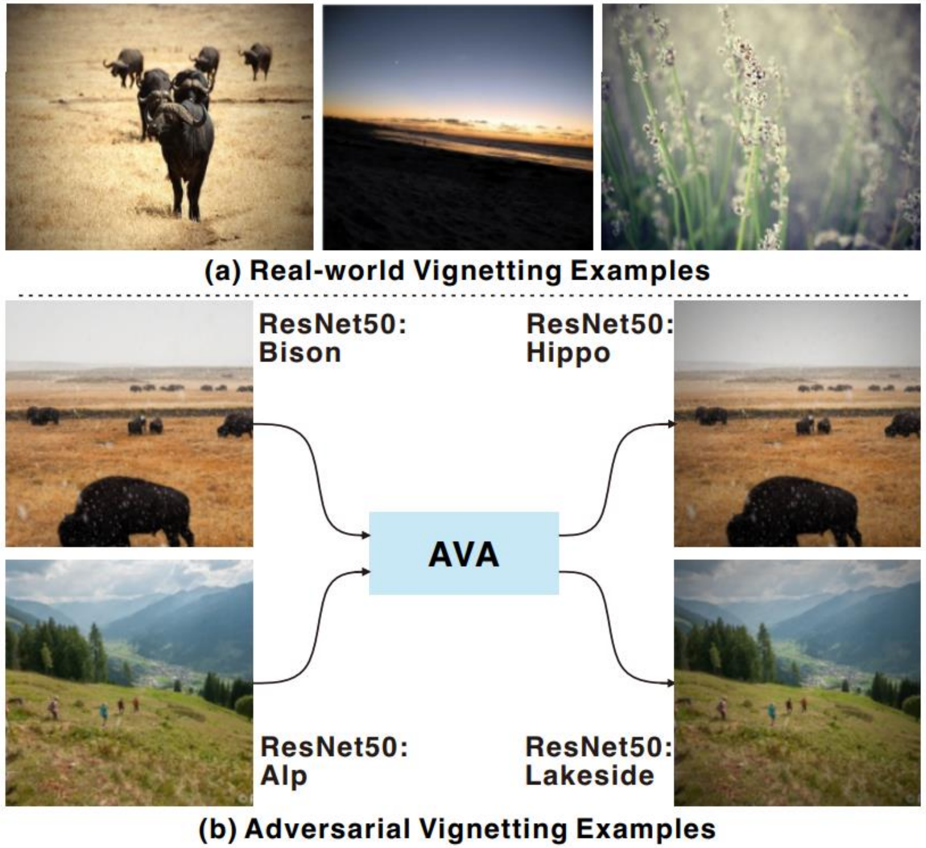
Attack

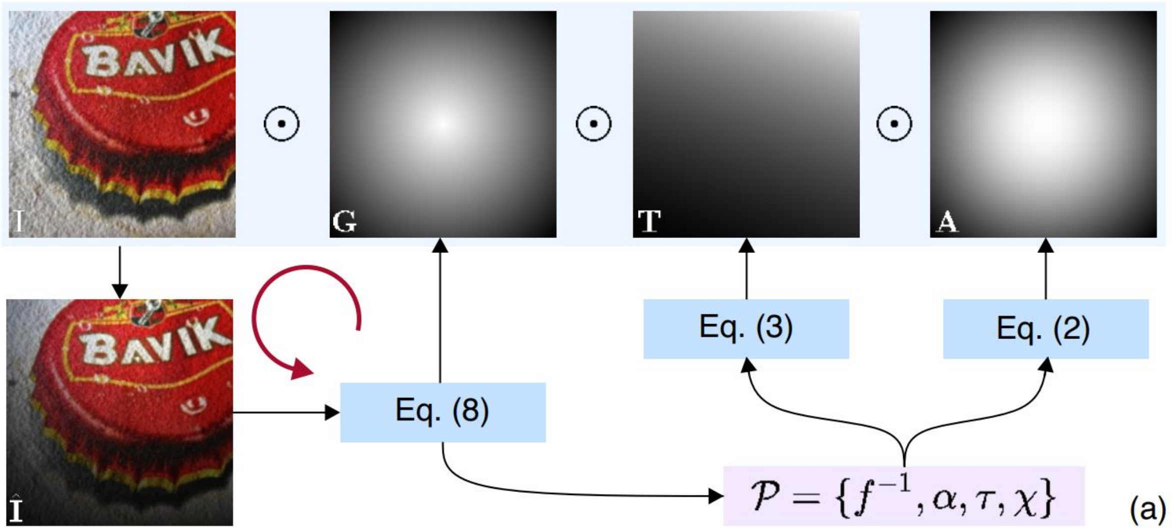
Figure II: Comparison between adversarial-blurred images and blurred images for training deblurring models.

Figure VII: Comparing the visualization examples of $ABBA_{physical}$ with those of ABBA.

Guo Q, Juefei-Xu F, Xie X, et al. Watch out! motion is blurring the vision of your deep neural networks[J]. NeurIPS, 2020

Attack





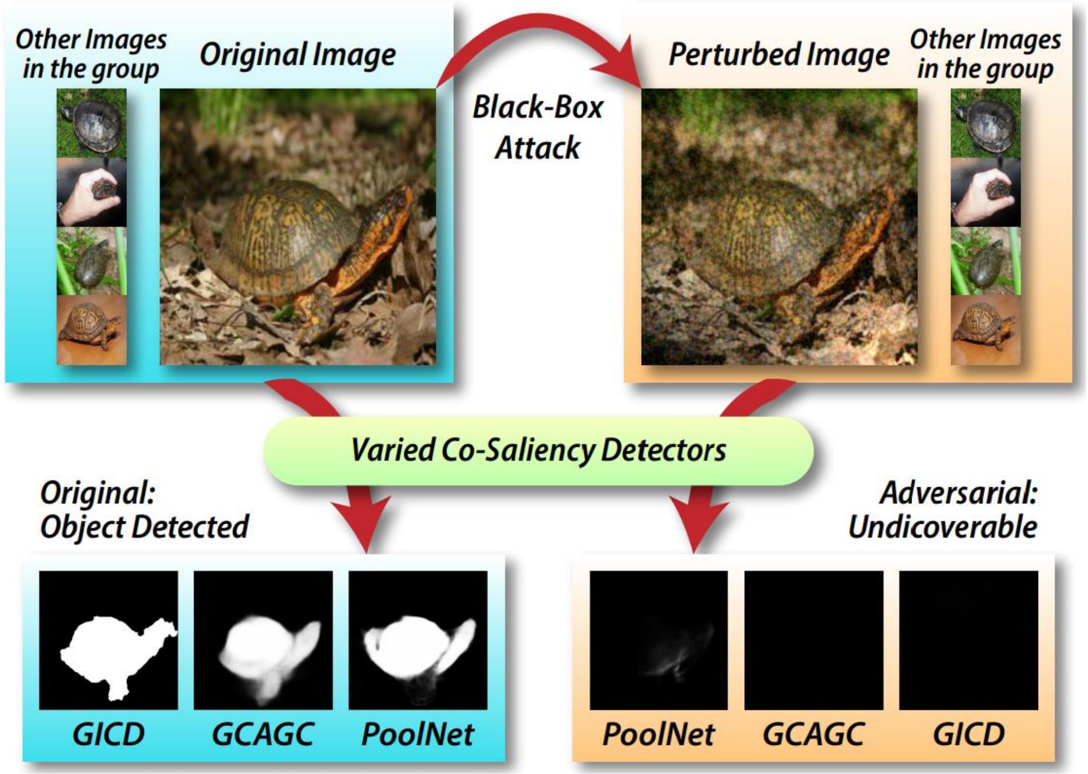
Tian B, Juefei-Xu F, Guo Q, et al. AVA: Adversarial vignetting attack against visual recognition[J]. IJCAI, 2021

Attack

Figure 4: Visualization results of different ball bound for f and α.

Tian B, Juefei-Xu F, Guo Q, et al. AVA: Adversarial vignetting attack against visual recognition[J]. IJCAI, 2021

Attack



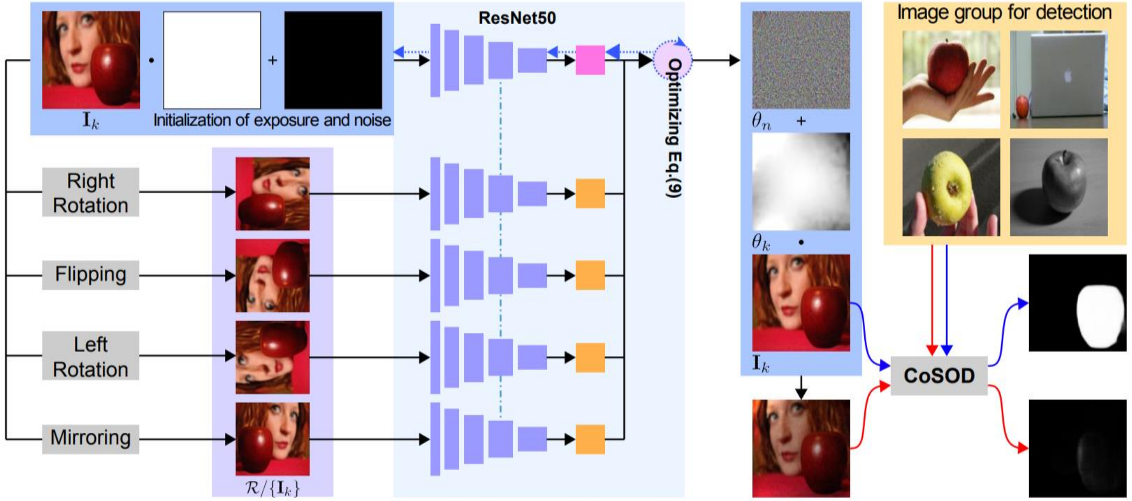
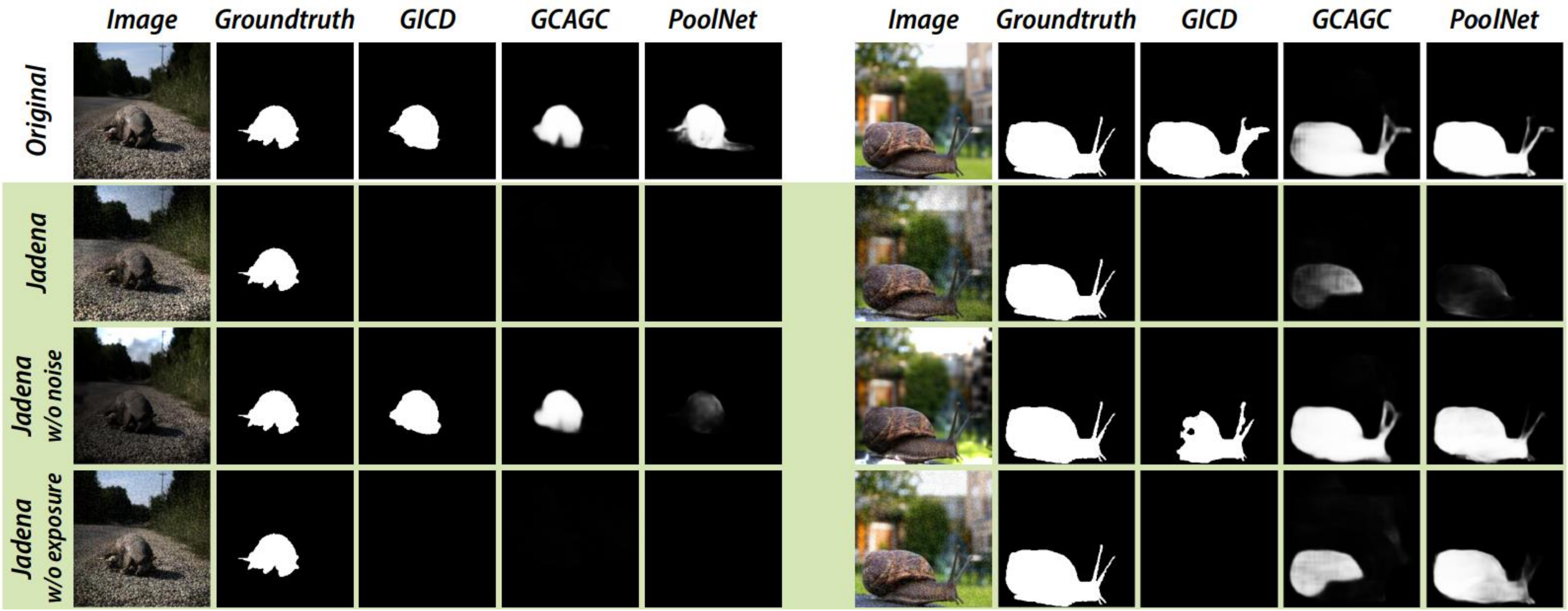


Figure 2. Pipeline of the *joint adversarial exposure and noise attack*. The clean image is augmented to generate references and the gradient back-propagates along the blue dashed lines.

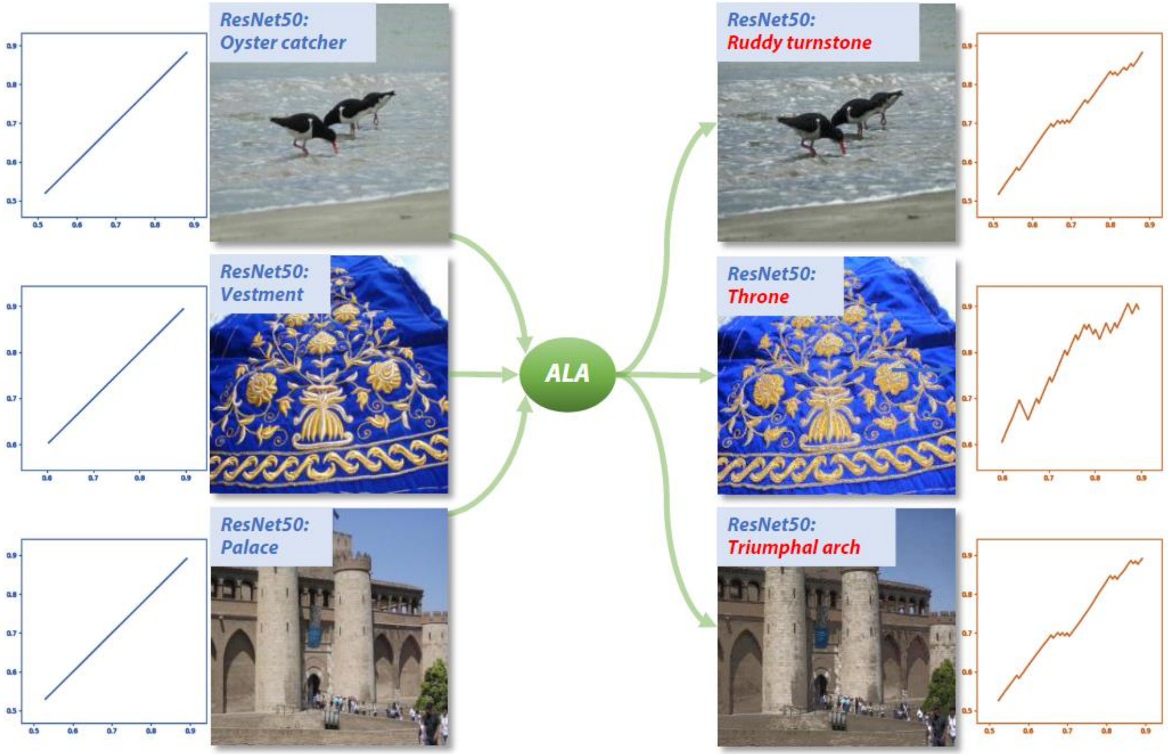
Gao R, Guo Q, Juefei-Xu F, et al. Can you spot the chameleon? adversarially camouflaging images from co-salient object detection[C], CVPR. 2022

Attack



Gao R, Guo Q, Juefei-Xu F, et al. Can you spot the chameleon? adversarially camouflaging images from co-salient object detection[C], CVPR. 2022

Attack



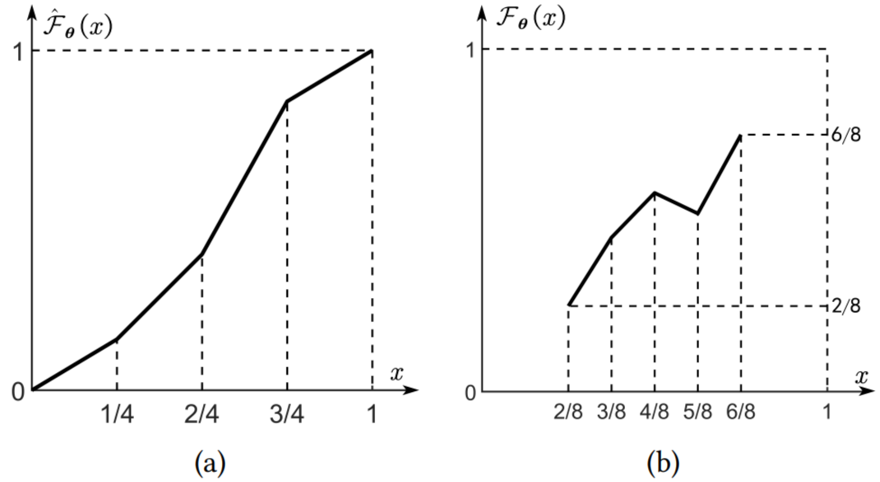
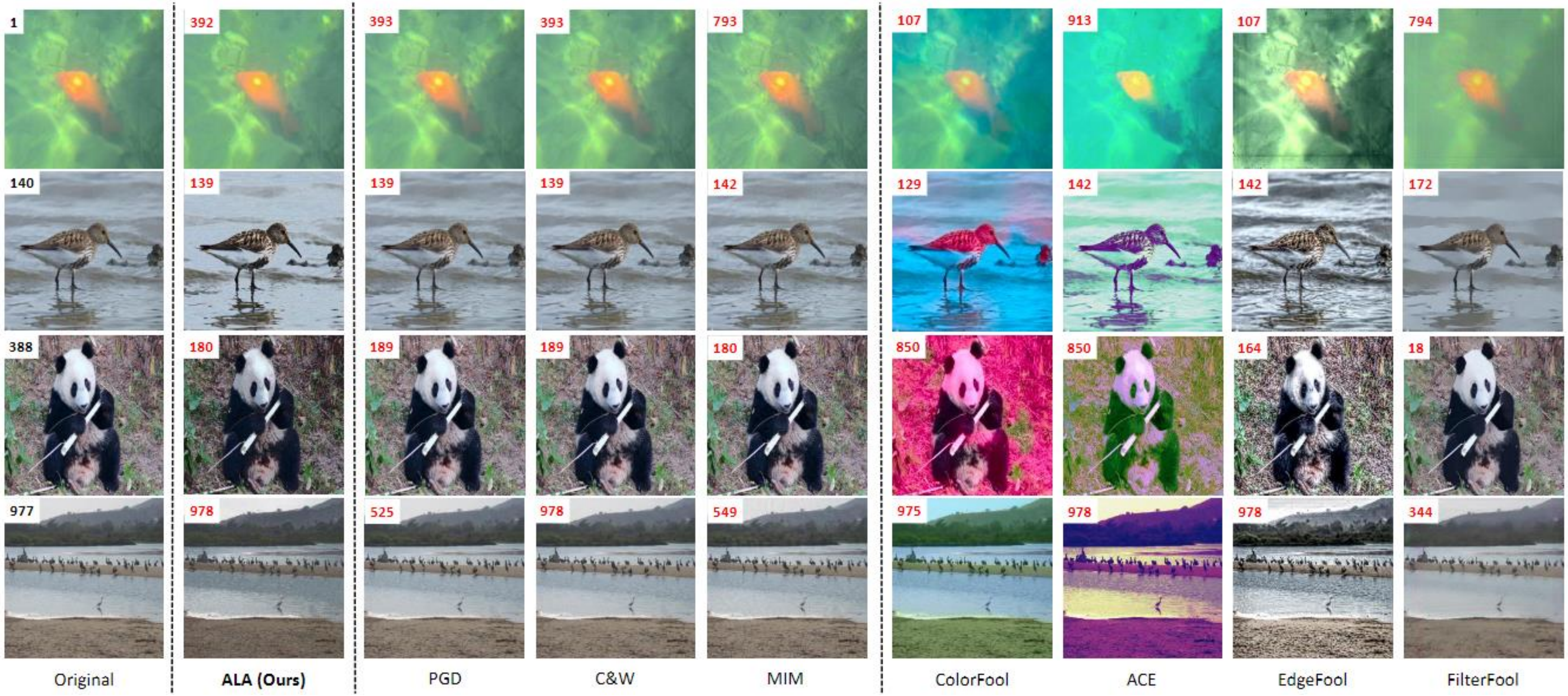


Figure 2: (a) monotonic filter $\hat{\mathcal{F}}_{\theta}$. (b) scene-adaptive filter $\mathcal{F}_{\theta}$ with the valid range from 2/8 to 6/8. Both filters are segmented into 4 pieces, *i.e.*, $T = 4$ in Eq. (1).

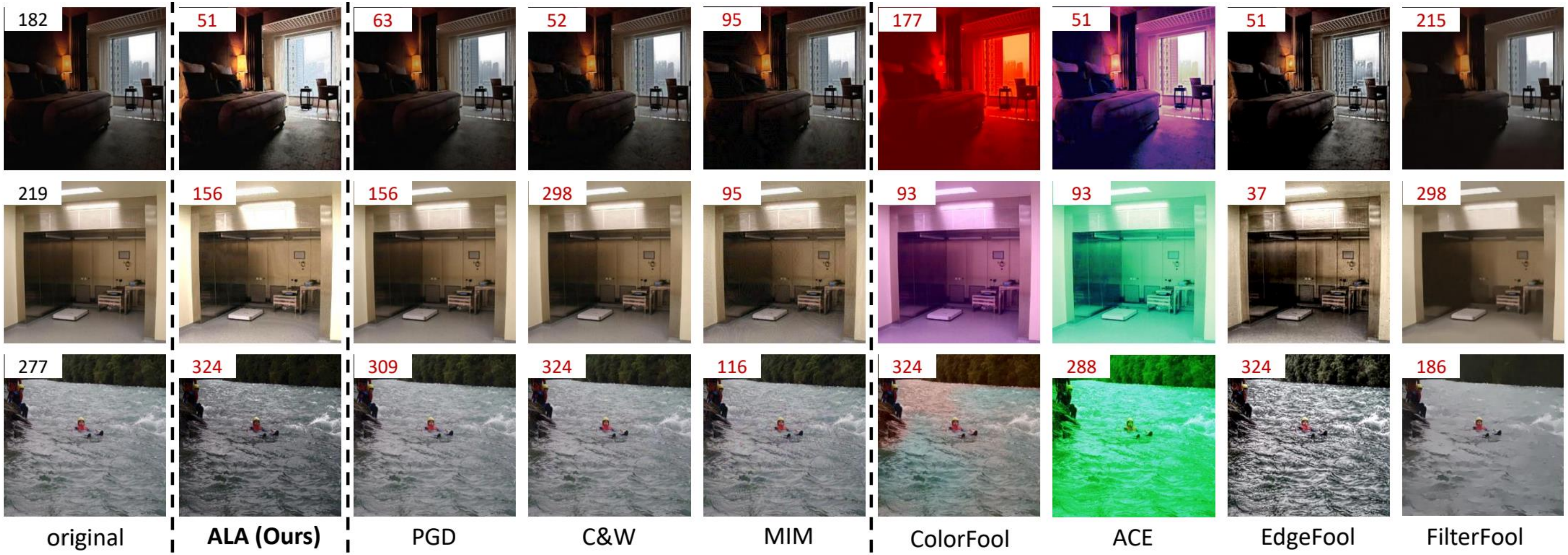
Huang Y, Sun L, et al. ALA: Adversarial lightness attack via naturalness-aware regularizations[C]. ACM MM, 2023.

Attack



Huang Y, Sun L, et al. ALA: Adversarial lightness attack via naturalness-aware regularizations[C]. ACM MM, 2023.

Attack



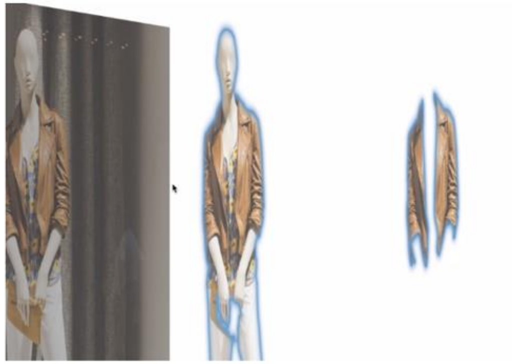
Huang Y, Sun L, et al. ALA: Adversarial lightness attack via naturalness-aware regularizations[C]. ACM MM, 2023.

Large Model

- **Segment anything model (SAM)**
  - ➢ **universality of perception system**

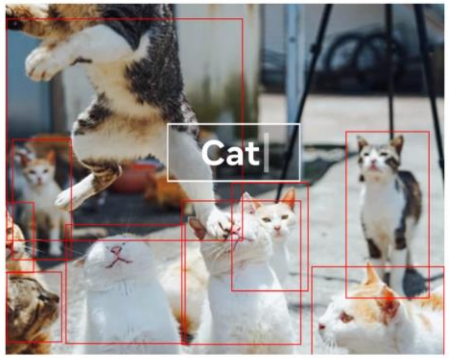
Prompt it with interactive points and boxes.

Automatically segment everything in an image.



Generate multiple valid masks for ambiguous prompts.

SAM can take input prompts from other systems, such as in the future taking a user's gaze from an AR/VR headset to select an object. This footage uses our **open sourced Aria pilot dataset**.

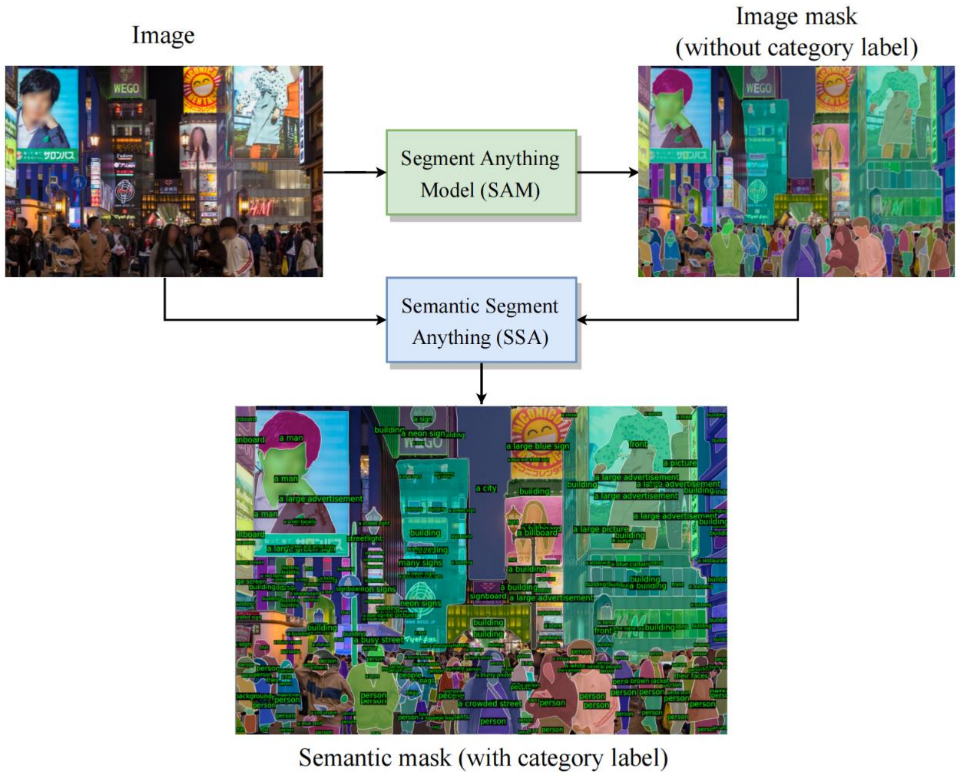


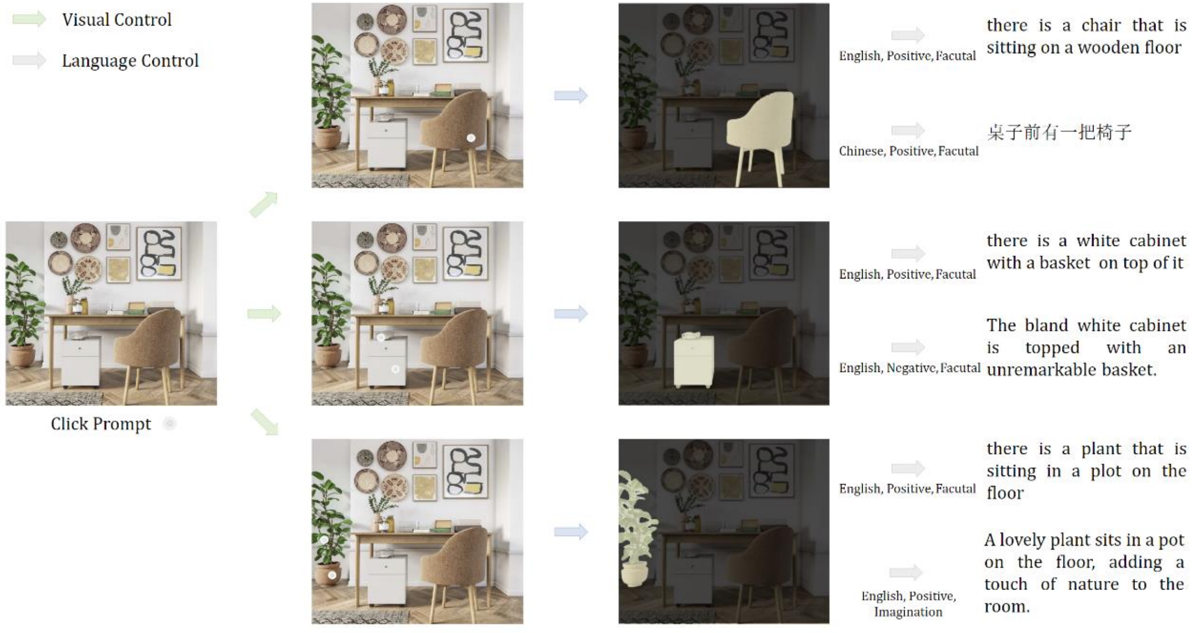
Bounding box prompts from an object detector can enable text-to-object segmentation.

https://segment-anything.com/

Large Model

- **SAM-related system**

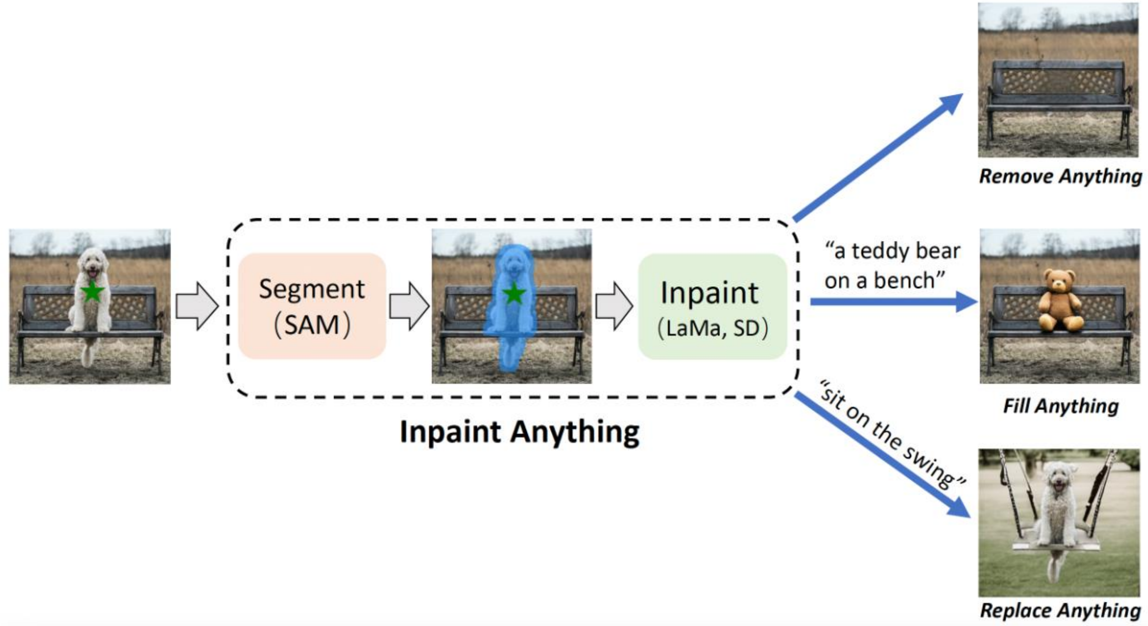


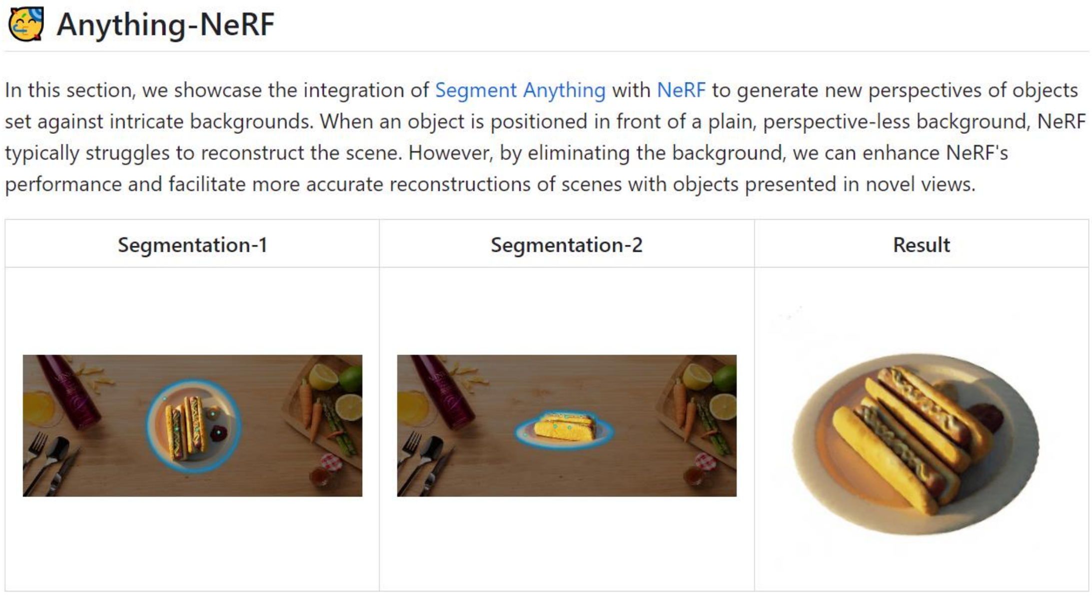


https://github.com/fudan-zvg/Semantic-Segment-Anything

https://github.com/ttengwang/Caption-Anything

- **SAM-related system**





Anything-NeRF

In this section, we showcase the integration of Segment Anything with NeRF to generate new perspectives of objects set against intricate backgrounds. When an object is positioned in front of a plain, perspective-less background, NeRF typically struggles to reconstruct the scene. However, by eliminating the background, we can enhance NeRF's performance and facilitate more accurate reconstructions of scenes with objects presented in novel views.

Segmentation-1	Segmentation-2	Result

https://github.com/geekyutao/Inpaint-Anything

https://github.com/Anything-of-anything/Anything-3D

Evaluation

- **Adversarial attack**

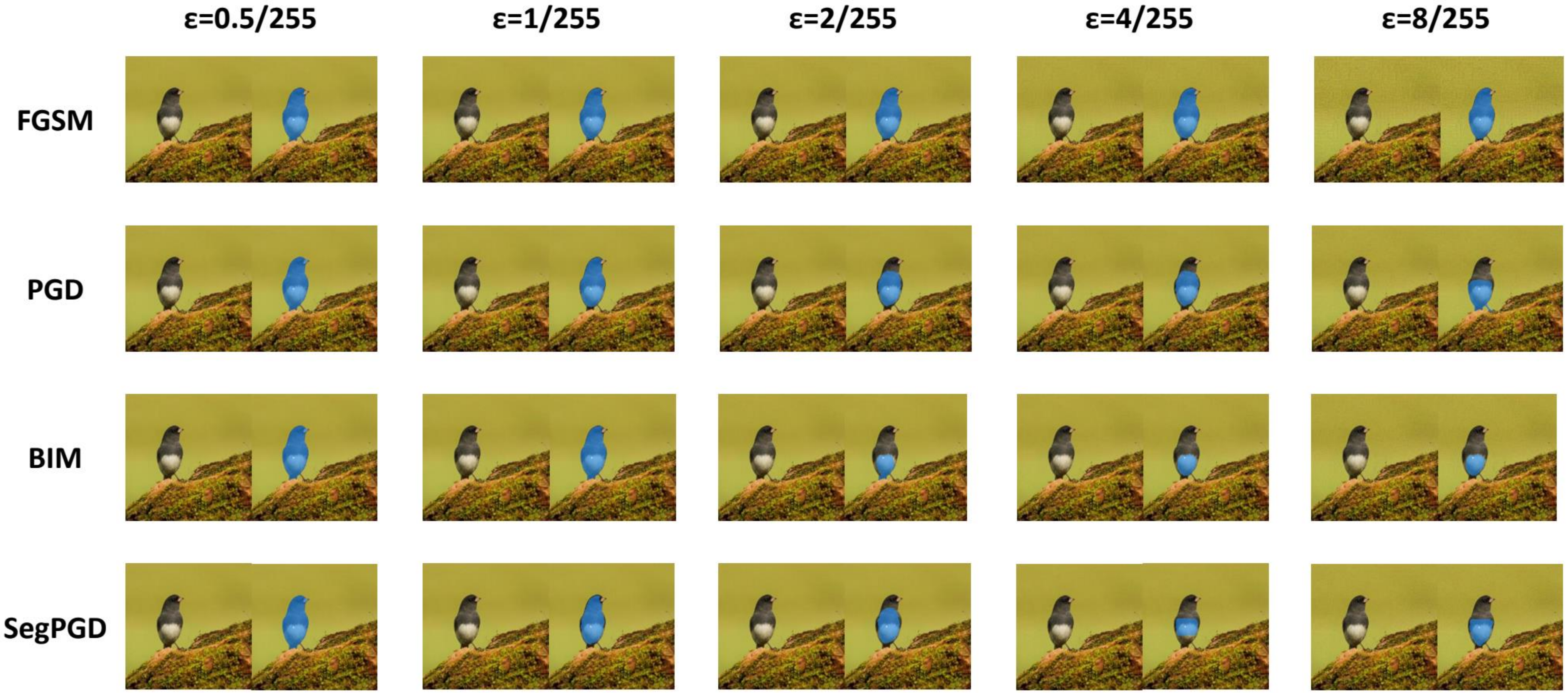


Fig. 1: Adversarial attacks examples under 4 kinds of attacks with 5 different severities and corresponding masks predicted by SAM.

Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

- **Corruptions**

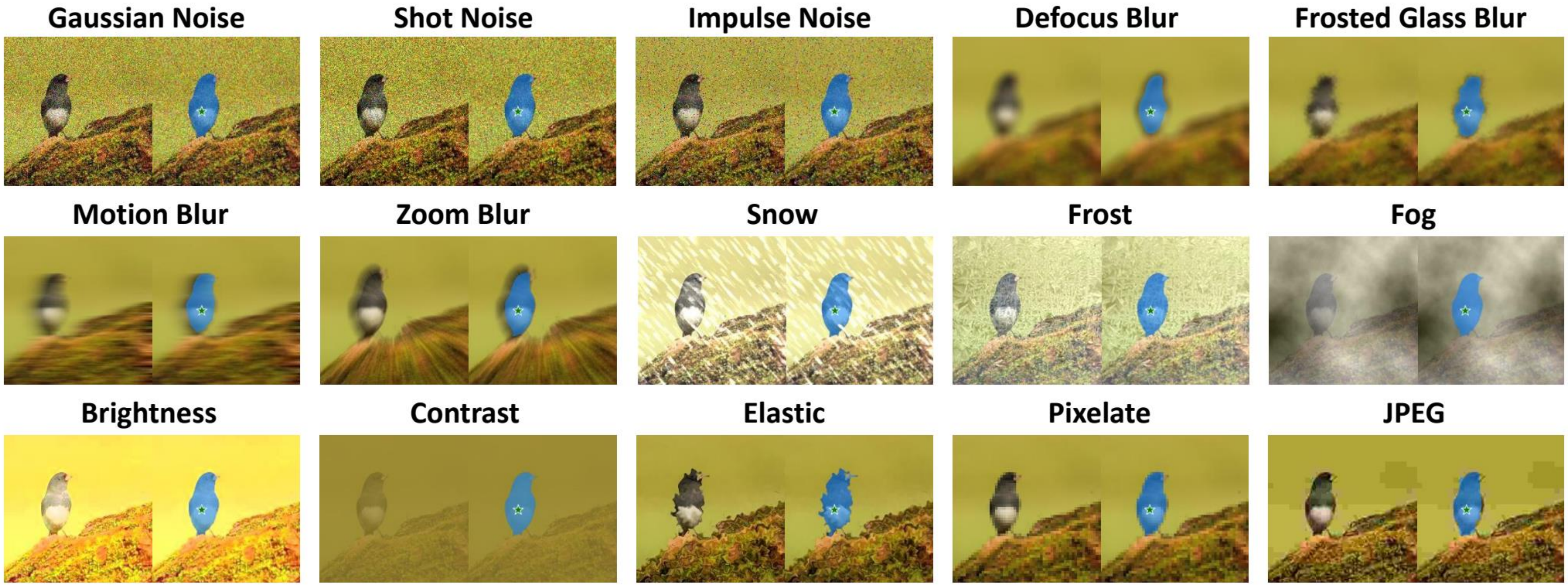


Fig. 2: Corruption examples of 15 diverse types and corresponding masks predicted by SAM.

Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

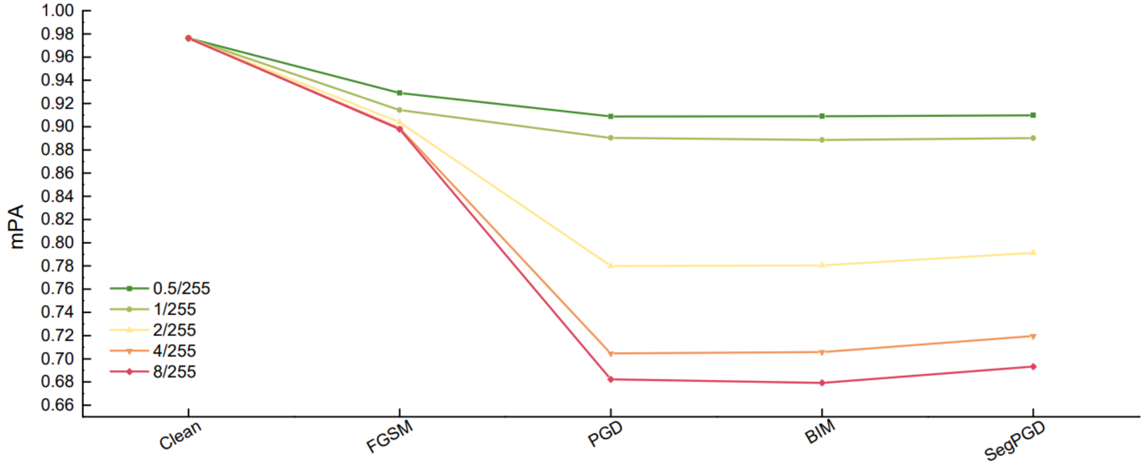


Fig. 3: The mPA values of SAM on SA-1B under 4 adversarial attacks and 5 severities.

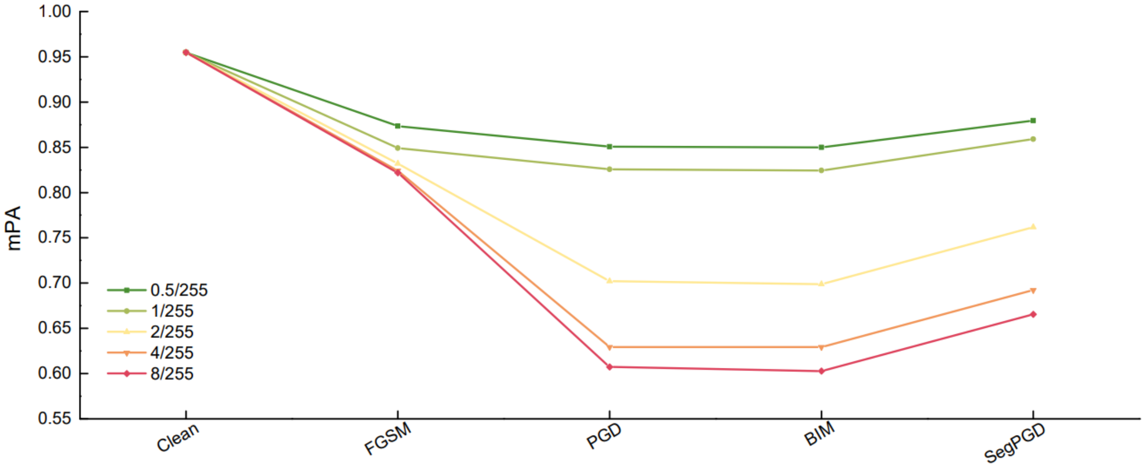


Fig. 4: The mIoU values of SAM on SA-1B under 4 adversarial attacks and 5 severities.

Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

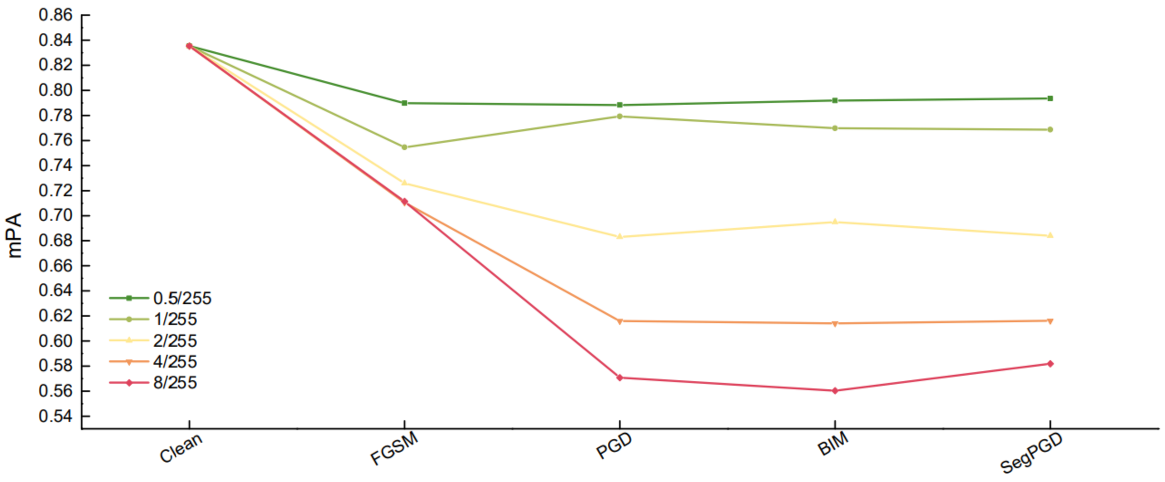


Fig. 11: The mPA values of SAM on KITTI under 4 adversarial attacks and 5 severities using MSE loss.

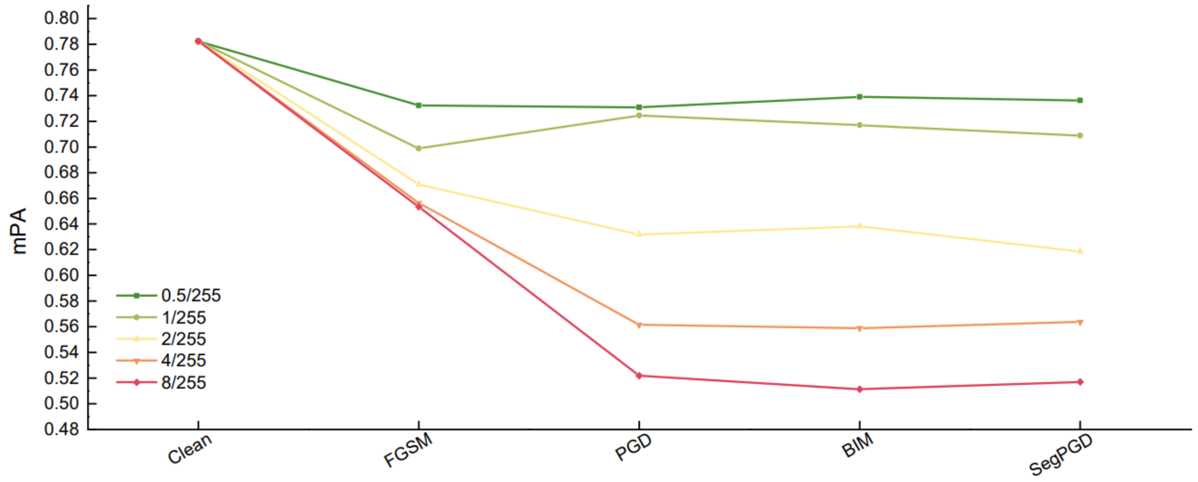


Fig. 12: The mIoU values of SAM on KITTI under 4 adversarial attacks and 5 severities using MSE loss.

Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

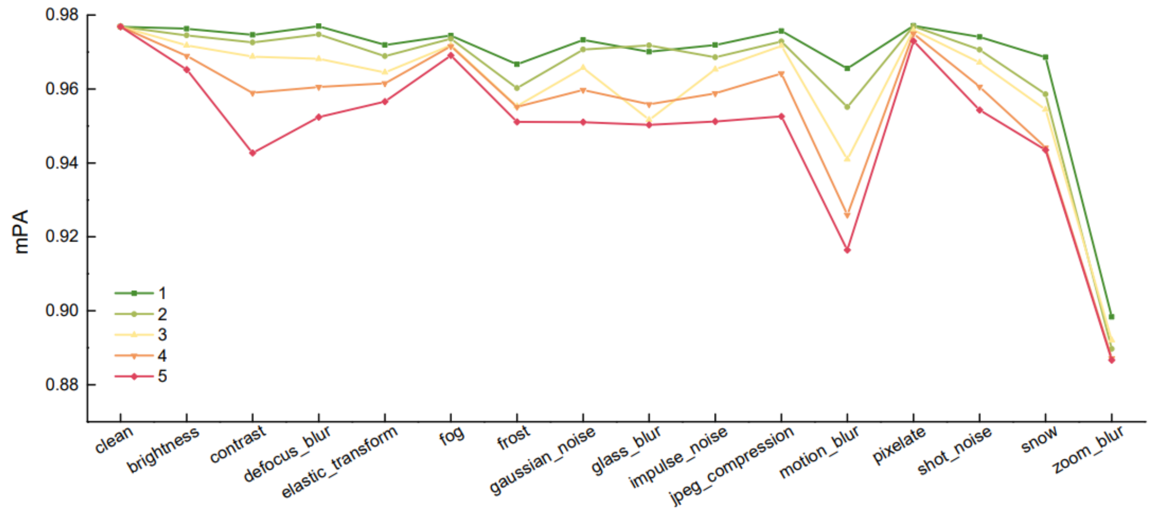


Fig. 17: The mPA values of SAM on SA-1B under 15 corruptions and 5 severities.

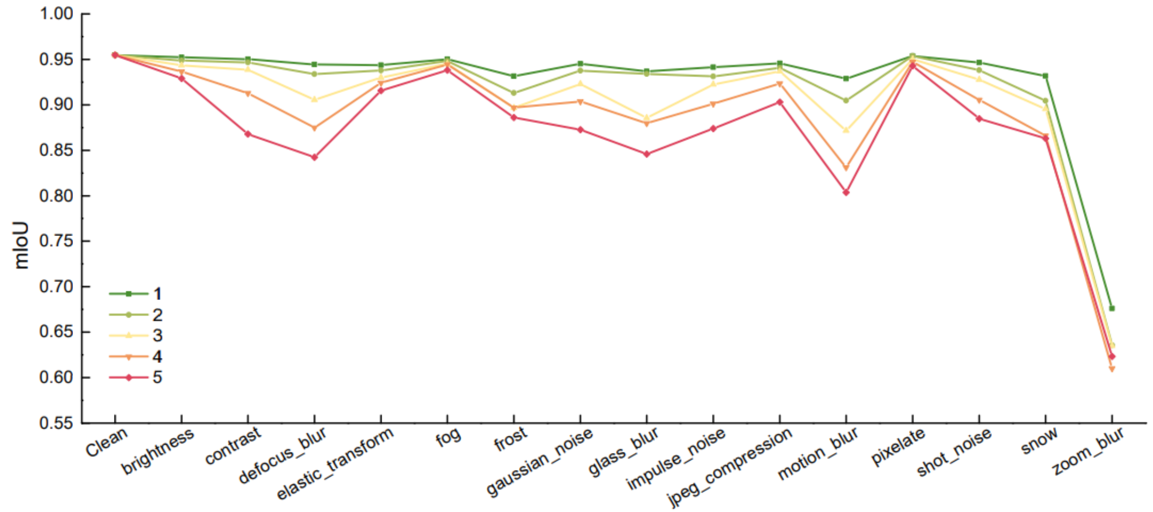


Fig. 18: The mIoU values of SAM on SA-1B under 15 corruptions and 5 severities.

Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

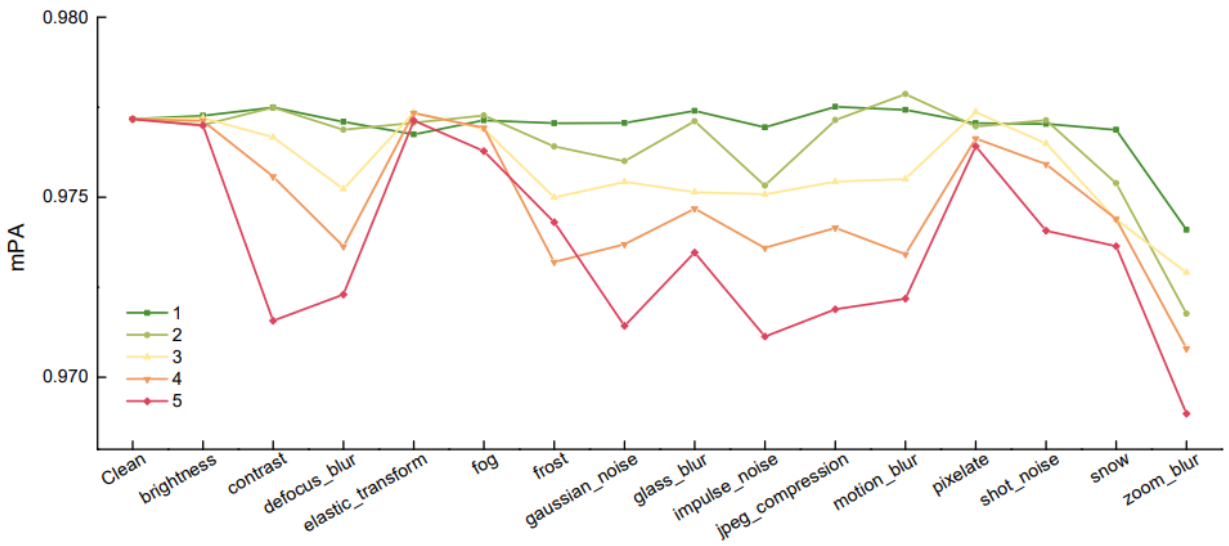


Fig. 19: The mPA values of SAM on KITTI under 15 corruptions and 5 severities.

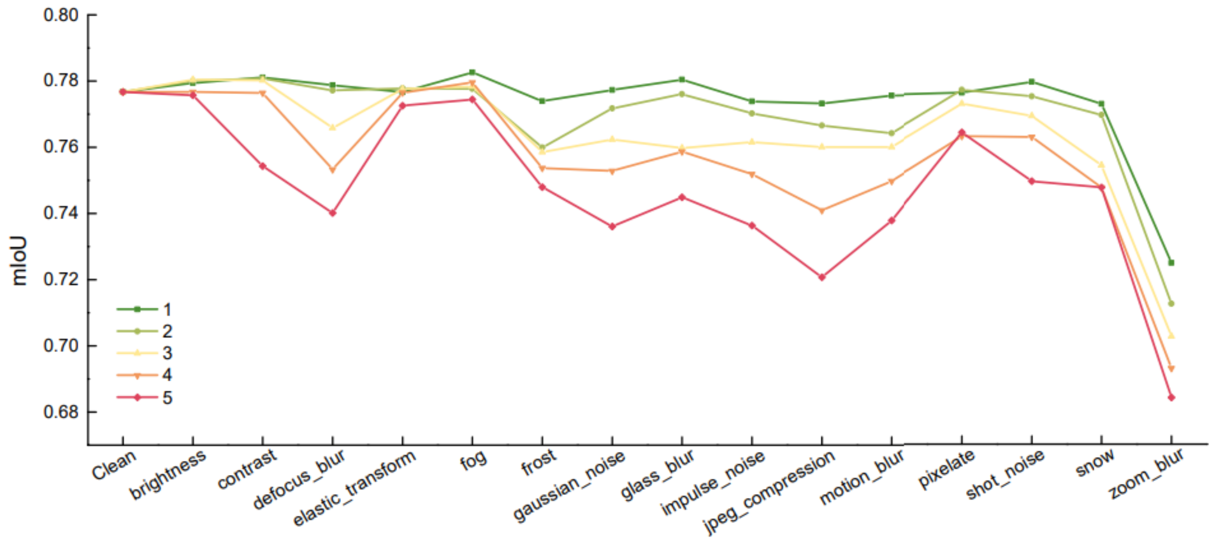
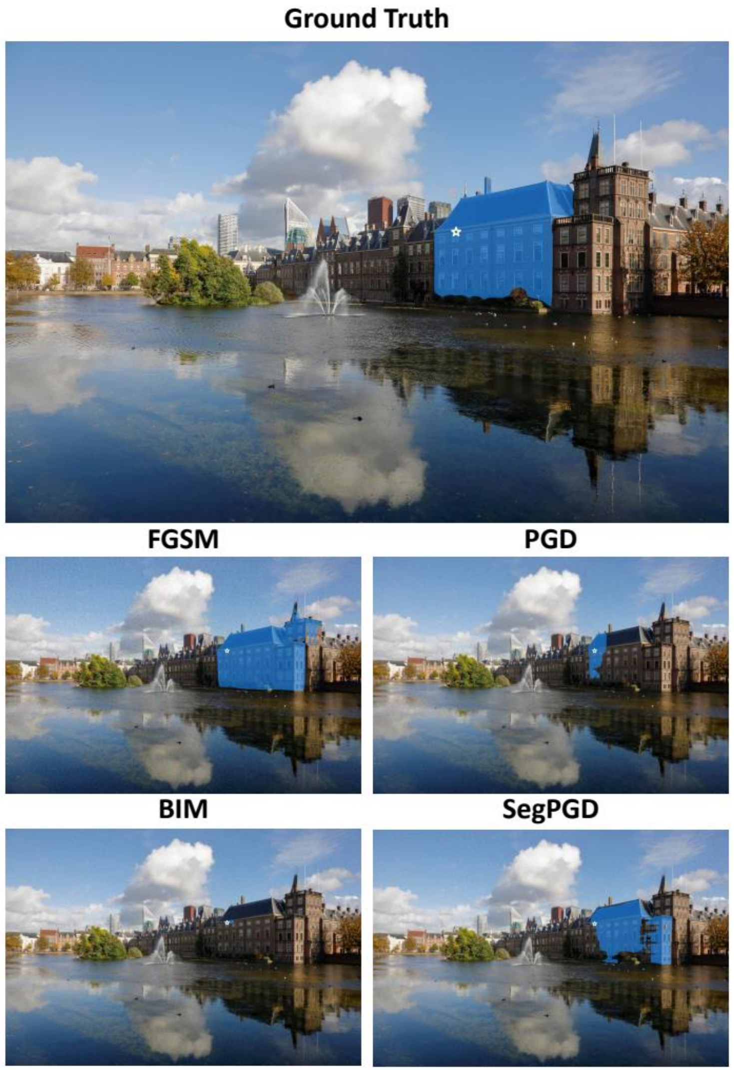


Fig. 22: The mIoU values of SAM on big objects of KITTI under 15 corruptions and 5 severities.

Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

- **SA-1B**



Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

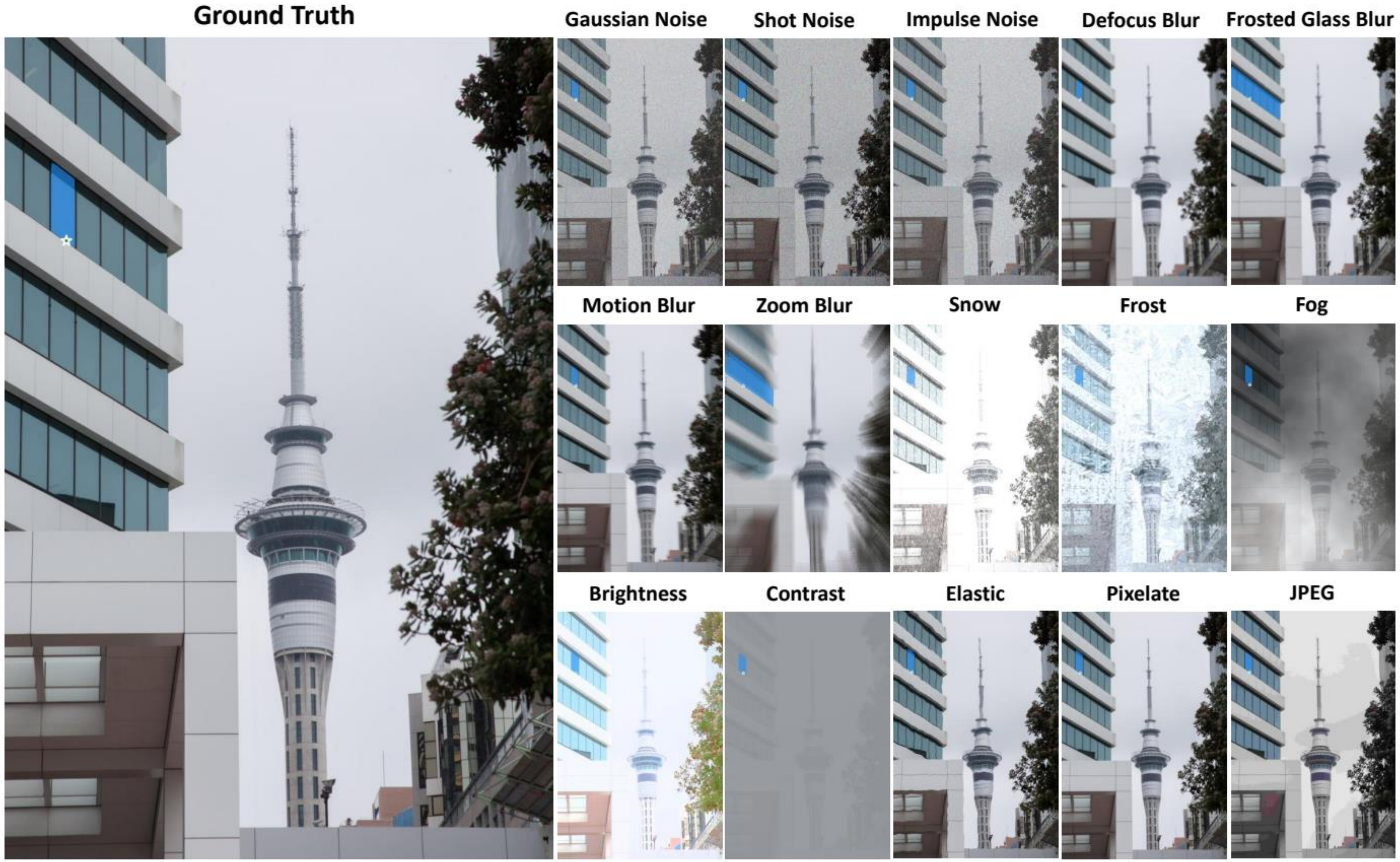
Evaluation

- **KITTI**

Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

- **SA-1B**



Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

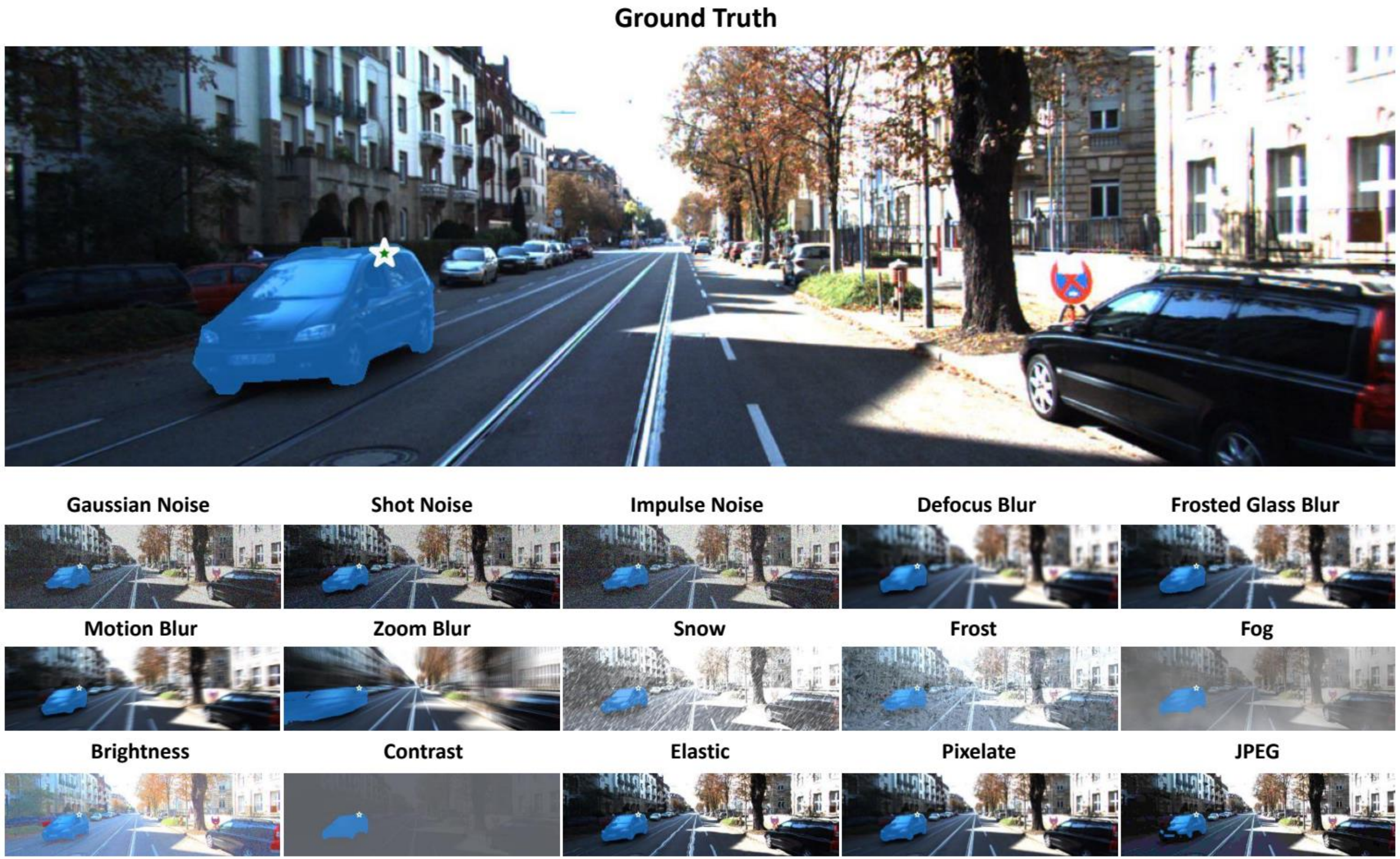
Evaluation

- **SA-1B**

Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

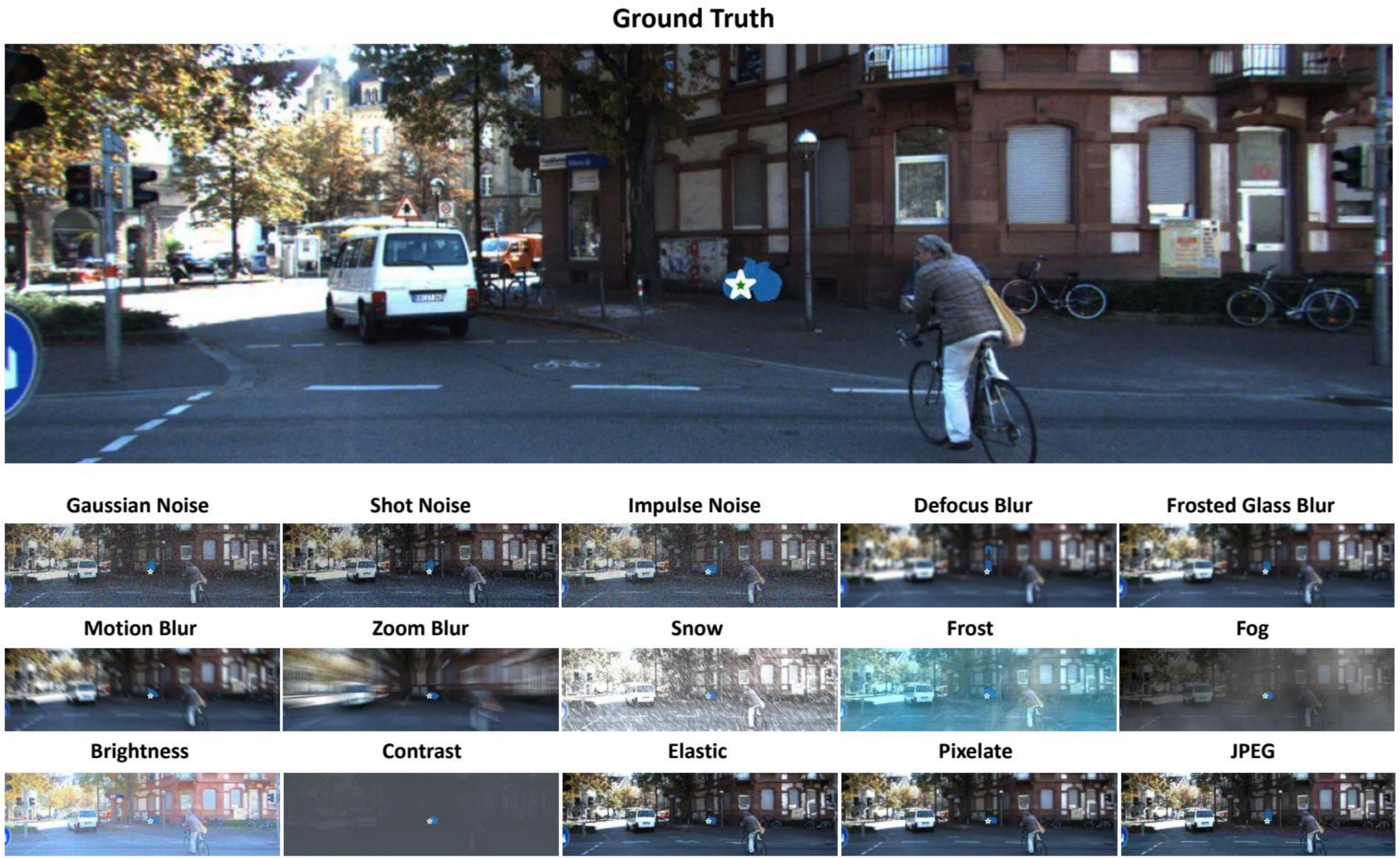
- **KITTI**



Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Evaluation

- **KITTI**



Huang Y, Cao Y, Li T, et al. On the robustness of segment anything[J]. arXiv

Thanks for listening!